

Durham Research Online

Deposited in DRO:

23 November 2018

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Bergamin, Joshua A. (2018) 'An excess of meaning : conceptual over-interpretation in confabulation and schizophrenia.', *Topoi.* .

Further information on publisher's website:

<https://doi.org/10.1007/s11245-018-9609-6>

Publisher's copyright statement:

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:


- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.



An Excess of Meaning: Conceptual Over-Interpretation in Confabulation and Schizophrenia

Joshua A. Bergamin¹ 

© The Author(s) 2018

Abstract

I argue that ordinary (non-pathological) confabulation is a side-effect of an interpretive faculty that makes sense of the world by rationalising our experience within the context of a personal and cultural narrative. However, I argue that a hyperactivity of the same process manifests as schizotypy—latent schizophrenic tendencies—that can lead to extreme dissociation of interpretation from experience. I first give a phenomenological account of the process of interpretation, arguing that it is enacted through the creation of conceptual cognitive content from an originary non-conceptual experience. I then ground this account in empirical evidence, showing how our direct perception is moulded and adapted as it is conceptualised to fit our individual and collective expectations. Finally, I argue that hyperactive over-interpretation results in schizotypic dissociation, thus suggesting that schizophrenia should be understood as a disorder of interpretation, the extreme end of a spectrum that includes ordinary confabulation.

Keywords Confabulation · Schizophrenia · Conceptual content · Interpretation · Narrative self

1 Introduction

In this contribution to the special issue on confabulation, I argue that the human mind is an ‘interpretive organ’ that makes sense of the world and forms our identity by interpreting our experience within the context of a narrative developed through our personal and cultural history. This ongoing process involves an inferential rationalisation that often leaves gaps between the source of our actions and what we interpret ourselves as doing. Such gaps are frequently filled by *confabulation*, the unconscious creation of an objectively-false but sincerely-believed narrative that attributes first-person agency and ownership to unconsciously-initiated actions. While much of the literature on confabulation focuses on pathological cases, narrative-generation more broadly is essential to our self-identity and I argue that the process of interpretation—including confabulation—plays a key role in maintaining robust mental health. However, a hyperactivity of the same interpretive process manifests as schizotypy—latent schizophrenic tendencies—and can

lead to an extreme dissociation of interpretation from experience. The study of confabulation, therefore, offers important clues in understanding the fine balance that storytelling and identity-creation maintain in the spectrum of mental health.

I argue this over four sections. In the first two sections, I give a theoretical background of the interpretive faculty and its relation to originary experience, arguing that confabulation involves a conceptual mental content that is created from a more originary, non-conceptual experiential content. Drawing on a synthesis of the arguments made by Dreyfus and McDowell in their debate over the role of mental content in action, I argue that the interpretive act that underpins both narrative-creation and confabulation involves a *translation* of the non-conceptual content of our direct experience into the propositional, conceptual content of reflective consciousness. Such narrative-creation is itself, therefore, unconscious, as this very process *creates* the first-person subject of self-consciousness. I then ground this account in empirical evidence, using examples from psychology and neuroscience to show how our direct, unconscious modular perception is moulded and adapted as it is conceptualised to fit our individual and collective narratives. In the final section, I argue that a hyperactive form of this process leads to the dissociation that is the hallmark of schizophrenia, where an individual’s interpretation diverges both from its

✉ Joshua A. Bergamin
joshua.bergamin@uqconnect.edu.au

¹ Department of Philosophy, University of Durham, 50-1 Old Elvet, Durham DH1 3HN, UK

objective source and from the collective cultural narrative. I close with some thoughts on future applications of work on confabulation and interpretation for mental and societal health.

2 Concepts, Coping, and Subjectivity

Human beings, according to a long tradition, are rational animals. We act for *reasons*, and common wisdom considers these to be transparent (at least to ourselves), and to connect our actions with our beliefs about the world, ourselves, and our intentions. However, the phenomenon of confabulation shows that the reasons we express for an action do not always align with what stimulates us to act. In this first section, I will argue that our rationality has less to do with acting for concrete reasons, which is something we share with other animals, but rather that *rationality* describes a form of cognition that has a conceptual, propositional content, one that is not shared with other animals, nor with our own spontaneous enactment of practised skills. I will argue that these two forms of content align with, and support, a model of the mind that posits two, interacting cognitive systems: a non-rational faculty that enacts our spontaneous as well as automatic moments of agency, contrasted with a deliberative, rational faculty which is the site of our conceptual cognition, and which interprets both the world and our own actions. A mismatch between the content of a subject's beliefs and the intentional content of their actions, I will suggest, is the source of confabulation, both in normal, non-pathological cases, as well as in the more extreme cases resulting from brain injury or mental disorder.¹

These two faculties, and their differing content, lie at the centre of Hubert Dreyfus' debate with John McDowell (Dreyfus 2005, 2007a, b, 2013; McDowell 2007a, b, 2013). In this exchange, Dreyfus (2007a, p. 352) argues that our 'smooth coping'—our pre-reflective embodied acts of agency—is "non-mental... non-conceptual, non-propositional, non-rational and non-linguistic," and must

therefore be contrasted against what has traditionally been termed 'thinking,' that is, with formulating and deliberating on conceptual propositions. McDowell (2007a, p. 338), on the other hand, maintains that our basic experience is already conceptually articulated, in the sense that the content of our coping experience would undergo no transformation if we were to express it propositionally. That is, although we might not always explicitly articulate propositions, our actions and experience are structured such that they *could* be so expressed.

The debate is complicated by the centrality of the terms 'conceptual' and 'non-conceptual,' which are the subject of ongoing debate in different philosophical traditions, with little consensus even on the nature of 'concepts' themselves. To understand the debate between Dreyfus and McDowell, it is important to appreciate what sense of conceptuality is at stake in coping, particularly over the extent that concepts need to be general, public, and communicable—that is, propositional. Taylor (2002, p. 11) argues that we need some 'minimal sense' of conceptuality that is none of these things whenever we talk of an agent pursuing an activity, since coping "can't be understood in just inanimate-causal terms." At a bare minimum, any animal's survival presupposes a capacity to 'pick out' what is relevant from what is not, although whether we ought to call this conceptual or not is a question of its own. But even if we were to call such minimal 'picking-out' conceptual, it would not be the form of conceptuality that is of real interest in the debate, what Rietveld (2010, p. 190) calls "strong rationality," which should be distinguished from anything we ascribe even to higher animals. 'Strong' conceptual capacities are "dependent on language-acquisition" and belong to a "linguistic or reflective faculty" (*ibid.*).

'Non-conceptual,' as Dreyfus in particular uses it, therefore diverges from certain analytic discussions of non-conceptuality. Carruthers (2011, p. 48), for example, uses 'non-conceptual' to refer to sensory experience so fine-grained as to be inexpressible (such as the different shades of green in a meadow), but not the experienced green entities (the grasses, trees, shrubs) themselves. Saying, therefore, that animals lack (strong) concepts is not to say that they can't experience trees or shrubs, nor does it imply that non-conceptual coping involves no experience of entities. Indeed, Dreyfus follows Heidegger in holding that sensory experience is normally always bound to the experience of a *thing*: "What we 'first' hear is never noises or complexes of sounds, but the creaking waggon, the motor-cycle" (Heidegger 1962, p. 207). In the context of the debate, then, 'coping' aligns with the agency of what Zahavi (2005, p. 21), following Sartre, calls "pre-reflective consciousness," in contrast to the "reflective" consciousness whose use of propositional concepts is uncontroversial.

¹ There is an ongoing debate over whether perception also has a conceptual content (see Crane 2013). However, even in cases of pathological confabulation—such as when a patient does not believe their hand to be their own (Sandifer 1946, pp. 122–123)—the perception (of the hand) seems unaffected; it is the subject's *belief* (that it is not their hand) that is at issue. One might argue that the issue is with the patient's proprioceptive rather than visual perception, but Hirstein (2005, pp. 10–11) notes that patients with body dissociation continue to confabulate even when forced to admit that they do not have normal sensation or control over their body parts. Hence, throughout this paper I will be concerned rather with the conceptual content of belief, and what I will argue is the non-conceptual content of agency, while remaining open to other debates on this very complex topic (see Siegel 2016 for review).

The debate, then, can be seen as over the extent and the manner in which our smooth-coping draws upon ‘strong’ reflective or propositional concepts, and centres on the question of whether embodied agency can be understood independently of conceptual rationality (McDowell 2007a, pp. 344–345). Dreyfus (2007a, p. 354) believes it can, and most of his arguments are rooted in the observation that experts in various fields—baseball, chess, music—tend to perform at their best when they are not (reflectively) thinking about what they are doing. This, he argues, shows that concepts, which are evidently involved in *thinking*, must be absent in coping. One of Dreyfus’ favourite arguments is based on the case of baseballer Chuck Knoblauch, a major league second-baseman who became (in)famous after falling into a curious pattern of throwing errors. He would repeatedly make mistakes with simple throws, made over short distances when he had ample time ahead of him. Yet perhaps even more strangely, during the same period he remained capable of accomplishing very difficult throws under time-pressure, with the high level of skill and accuracy that got him into the big leagues in the first place.

Dreyfus concludes from this story that Knoblauch’s problem was that, on the easy throws, he was *thinking*—that is, employing strong, abstract concepts. Rather than trusting his body to reflexively aim and complete the throw, Knoblauch became entangled in reflective thought, facing the concept-heavy problem of calculating how hard and at what angle he should throw the ball he was holding, with the result that his throws were far less than perfect. During the high-pressure, complex throws—with no time to think—Knoblauch simply reacted with the refined, smooth-coping that he had formerly employed even on his simpler throws.

Extending this picture has many implications, including, as we will see, for the question of confabulation. Dreyfus’ assertion is that both our most expert as well as our automatic, everyday actions occur without the linguistic, conceptual mode of cognition that modern philosophy has long taken to be the mark of human mindedness (in Descartes’ ‘*cogito*’ or Kant’s ‘*Ich denke...*’ for example). Our coping utilises cognitive capacities that we share with non-human animals. This is in sharp contrast to McDowell who emphasises the difference between rational and non-rational animals. For while it is tempting to equate our smooth-coping actions with the behaviour of animals—as a fielder chasing a ball seems to act with the same smoothness and immediacy as a cat chasing a mouse—humans nevertheless *are* the only animals that can play baseball, because they are the only animal capable of learning the concepts of the equipment and the rules.

Of course, Dreyfus does not deny this. Rather, he wants to show not only the ways in which our experience parallels and diverges from that of animals, but more interestingly, how these two layers interact (Dreyfus 2007a, pp. 354–355).

Indeed, Dreyfus’ account of embodied coping is tied to his account of skill acquisition, where (with his brother Stuart) he describes how any skill—from chess to baseball to riding a bike—consists of a learning period in which performance is clumsy and thought-out, only following which we get the absorbed, reflexive coping of the expert in action (Dreyfus and Dreyfus 1986, pp. 19–36). From this, he argues that there is a change in intentional content as knowledge shifts from one layer to the other.

For example, most drivers will recall their awkward early attempts at trying to time pushing the clutch and the accelerator and shifting the gearstick, and the embarrassing stalls when this doesn’t go to plan. Novice drivers are often taught to employ rules of thumb, such as changing gear at a certain rate of RPM, or to associate a certain gear number with a certain speed. Such rules are helpful, yet remain imperfect, and as the driver improves they learn through trial-and-error to adjust the rules to specific situations, such as climbing a steep hill. All the same, even as the driver becomes competent, the phenomenological experience of the task is as intellectually ‘heavy.’ The driver must concentrate their attention on what they are doing, and novel situations may require explicit decision-making, an inner dialogue saying, for example, ‘this hill is really quite steep... I wonder if I should put it into second.’

For Dreyfus and Dreyfus, this sort of experience is contrasted against the highest stages of skill acquisition, which they call proficiency and expertise, where such explicit thinking plays a far smaller role. We can drive the winding country road while arguing with our passenger about whether our destination is before or after the bridge. As we climb the hill, we shift downgear automatically, our attention concerned with the amount of fuel left rather than the speed, RPM, or even the sound of the engine.

The phenomenological observation that explicit concepts—‘thinking’ and rules—fade as we become proficient at a skill is the other side of the observation that reflection, or the introduction of explicit thought, disrupts smooth coping. In essence, explicit thought reverts us back to the beginner level, just as seemed to happen with Knoblauch. In a similar way, we might be confidently and automatically shifting through the gears as we round corners and climb hills, yet if we’re suddenly asked to draw our attention to just how hard we push the clutch or exactly when we pull the gearstick, it is more likely that we will make an awkward mistake. Just as, when learning to drive, we necessarily employ a great deal of conceptual propositions to achieve a rather imperfect performance, so when we reflect, our otherwise smooth coping is ‘disrupted,’ and we no longer perform as an expert.

McDowell would object that, while our expert gear-change seems automatic, on attentive reflection we can attribute our downshift to an awareness of the car ‘struggling.’ But Dreyfus emphasises that this is not how the

coping is phenomenologically *experienced*, invoking Sartre's famous example of his experience chasing down a streetcar. "When I run after a tram..." says Sartre (2004, p. 13), "there is no *I*... I am then plunged into the world of objects... which present themselves with values, attractive and repulsive values, but as for *me*, I have disappeared." The street takes on a particular significance—the footpath, kerb, and other pedestrians appear as affordances or obstacles defined by the context of the action (cf. Gibson 1979, pp. 127–129), and indeed have a different significance than they would in a different context, such as, for instance, if we were trying to find the café where we are meeting a long-separated friend. However, Dreyfus (2007b, p. 373) points out with interest that in reflection, Sartre "can't help remembering himself" as the subject of his experience, when he actually does reflect backwards on it.

When Sartre says there is no 'I' in immersed experience, he is not claiming that the experience is not conscious, nor that it is not driven by an agent or self. Rather, what the phenomenology reveals is that prior to the experience of an 'I' is what Zahavi (2005, p. 106) calls a 'core' or "minimal self." This minimal self is "not something standing beyond or opposed to the stream of experiences but is rather a feature or function of its givenness" (*ibid*). While the 'I' of reflective consciousness is experienced as a subject among objects (and hence has a certain objectivity of its own), the 'core self' is neither subject nor object of consciousness, but an integral element of it. Zahavi (2005, p. 126) therefore suggests that, rather than considering this self the 'subject of experience,' we think of it as the "*subjectivity* of experience" [my emphasis]. This distinction between the *subjectivity* of coping and the *subject* of reflection will be important as we go along, though it should also be kept in mind that the subjectivity refers to the specific 'mine-ness' of coping, and doesn't imply that every aspect of coping is pre-reflectively conscious. As I will argue below, the phenomenon of confabulation shows that elements that are outside even minimal, pre-reflective consciousness can come to be interpreted in terms of the reflective ego.

3 The Rationalising Animal

It should already be becoming clear how the *post hoc* reflective rationalisation I have described relates to confabulation, which by definition involves the attribution of a proposition to an act of a subject. Before discussing confabulation in more depth, however, I want to argue that non-conceptual coping encompasses most of our actions, and that such actions are therefore non-rational—not in the sense that we don't have reasons for them, but that they do not involve propositional content. In an important sense, this mode of everyday cognition is something we share with non-human

animals, and is distinguishable from the exclusively-human conceptual cognition to which confabulation belongs.

It is important in this connection that we note Dreyfus' stress on 'everyday' expertise. Focusing overly on the complex actions of expert athletes or artists 'in the flow' seems to discount everyday activities such as running for a tram—activities that still require practice, but not necessarily in the same intensive way as high-level baseball. But such everyday activities are precisely the sort of thing that the theory of smooth-coping—in its Heideggerian origins—was developed to account for. A preoccupation with 'experts' leads us to forget that much of our daily life is actually lived (albeit 'inauthentically') in a flow-like state, and that what makes the activities of experts so impressive is their accomplishment of difficult tasks with the same non-rational reacting that most of us employ only on mundanities like crossing the street.

Following this line, Rouse (2013, p. 252) suggests that animals ought to be considered 'experts' who exemplify smooth-coping by reflexively accomplishing often-complex tasks with a minimum of planning. Most animals, to be sure, do not *learn* most of their tasks, at least not in the conceptually-loaded way that we learn the rules of baseball or how to change a tyre. Yet they do seem capable of incredibly intricate everyday tasks, making them look so simple as to appear automatic, but in reality employing a finely-tuned discrimination of the highest degree. A squirrel rustling through a canopy, for example, is not just running along, but taking in its surroundings through sight and touch, finding grips, avoiding obstacles, judging the gaps between twigs and branches, and leaping over the larger ones.

Schear (2013, pp. 285–286) therefore frames the Dreyfus-McDowell debate in terms of the 'Venerable Thesis'—*homo est animale rationale*: humans are to be understood in terms of their rationality, with rationality here defined by both Dreyfus and McDowell in terms of conceptual capacities. Schear spells out three possible outcomes. Firstly, it could be false; we might be contingently rational creatures who would still be recognisably human without a 'strong' rational capacity. Neither Dreyfus nor McDowell support this possibility. Secondly, a 'weak' form of the Thesis may be true. That is, rationality may be a central feature of the human, but it is only *one* of our capacities; no matter how important conceptual understanding is, there is a significant sphere of our life to which it does not apply. This seems to be Dreyfus' position. Thirdly, we find the 'strong' Venerable Thesis, that rationality "is the form of the human as such," in which rationality is understood to pervade and to be involved in all human activity, even that which does not seem to fit an understanding of rationality as 'reflective thought' or 'reason-grounded decision.' McDowell appears "committed" to this reading (*ibid*, p. 290).

McDowell (2007b, p. 369) illustrates this point by imagining playing frisbee with a dog in the park. A passer-by wanders through the game, notices the frisbee sailing in her direction and, spontaneously, reaches out and catches it—a perfect example of what Dreyfus would call smooth-coping. McDowell fully grants that there is no reflective thought involved in this scenario. If asked for a reason *why* she did what she did, he suggests, she would be lost for words. “No particular reason,” she might reply, “I just felt like it.” Nonetheless, McDowell maintains, her unreflective action was an exercise of her particularly human, rational capacities. Your dog, which might catch a frisbee just as spontaneously and effortlessly as our quick-reflexed new friend, would not be utilising any similarly rational capacities. “In the relevant sense,” says McDowell, “he has none” (*ibid*).

Although he doesn’t say as much, McDowell here makes an important distinction that is easily lost in discussions of reasons and rationality. In English (as in Latin) the root ‘reason’ (*ratio*) covers two concepts that are distinct in other languages. In German, for example, we find the word *Grund* for ‘reason why...’ (as in, our friend had ‘no reason’ for catching the frisbee), with *Vernunft* referring to the intellectual capacity. It is this second meaning of *ratio* that is at stake in the Venerable Thesis’ claim. As Okrent (2007, p. 110) points out, making a similar distinction between what he calls practical rationality and instrumental rationality, there is no controversy in granting animals *Gründe* or reasons *for* a certain behaviour—a wasp buries food *for* her unhatched babies, even if she could never be aware *of* that fact. Thus, even though we might poke around for a reason for the dog to catch a frisbee (to impress you? In hopes of a treat? For the sheer delight?), we have already seen with McDowell’s frisbee-catcher that the presence or absence of a clear *Grund* need not be connected to the presence of *Vernunft*.

It is *Vernunft*—conceptual rationality—that McDowell argues is pervasive in all human activity, and hence does not require the reflective deliberation that acting *for reasons* seems to imply. When McDowell says reason is pervasive, he does not mean that we actively weigh up *Gründe* for everything we do or see. Rather, he is talking about the capacity that makes deliberating on such *Gründe* possible in the first place. The faculty of *Vernunft* is pervasive, McDowell believes, because it can be applied to any act of the human agent, even if the agent herself cannot express any propositional content more than ‘I don’t know why I caught the frisbee; I just felt like it.’ That is to say, since we can *think* a content for any embodied act, such actions themselves must already be conceptually articulated. By the same reasoning, since a dog is not capable of rational thought, the same action on its part must be non-conceptual.

But there is another possibility—that the distinctively human trait is not a conceptual rationality that is present in and accompanies our unreflective actions; rather, *reflection*

itself is enacting that uniquely-human, rational (*vernünftig*) ability, which *creates* conceptual content out of the pre-conceptual. Carman (2013, p. 175) accuses McDowell of overlooking such reflective interpretation of content, and committing what he calls the ‘Scholastic Fallacy,’ the “illicit projection of the structure and content of reflection into unreflective experience.”² For Dreyfus, our basic coping is as momentary, reflexive and solicited as a dog’s, but we can interpret the content of that experience *conceptually* through our reflective faculty. Other animals, lacking this faculty, also thereby lack concepts.

Dreyfus (2007a, p. 360) locates in reflection an active process, a *transformation* of the pre-conceptual, embodied action into a conceptual thought. “Reflection *rationalises*” says Carman (2013, p. 166), emphasising that reflection is not a passive ‘looking’ at an already conceptual coping, but an active modification of primordial, embodied experience. When we reflect, we are not simply making a more explicit expression of the content of our involved smooth-coping. Rather, says Dreyfus (2007a, p. 360), “reflection must introduce some other content.” Reflection doesn’t just *discover* the implicit content; it *creates* it.

‘Create’ does not necessarily signify the act of a conscious ego-self; indeed, in several theories of consciousness (e.g., Dennett 1991; Baars 2002), what we call the subject or ‘ego’ might itself be said to be created *through* this reflective process. What is happening here is not a case of some homuncular self observing and making propositions about the content of memories, but a process of rationalisation that creates that very content, including the subject of the reflector, which is to say, the first-person. This subject or ‘narrative self’ is distinct from but dependent on the ‘minimal’ or ‘core’ self discussed above, which is an essential and inalienable component of the experiences around which the narrative self is constructed via acts of propositional reflection (Zahavi 2005, p. 105; Schechtman 2011, p. 398). Since the ego arises simultaneously as a feature (that is, the subject) of reflective propositional content, we—as that first person—are not usually aware of creating such content, which is experienced as though it were always there. Dennett (1991, p. 250) has argued that this storytelling is central to our sense of self, and while such stories generally derive from our direct experience, he observes that our conscious mind frequently ‘confabulates’ while constructing the

² O’Regan (2000) calls this kind of thinking the ‘refrigerator light illusion.’ Just as we should not conclude that the fridge light is always on because it seems to be every time we open the door to check, neither should we conclude that an experience is always conceptual (or conscious, etc.) because we encounter it this way each time we ‘check’ by reflecting. Cf. Dreyfus, above, on how Sartre ‘always finds’ the ego in reflection.

narratives that explain our actions, noting that such narratives feel convincing even when they are demonstrably false.

Gazzaniga's (1998a, b; Gazzaniga et al. 1977) work with split-brain patients also suggests that rational explanations come after the fact, and can be completely and unconsciously fabricated. In one experiment (Gazzaniga 1998a, p. 133), he tells the speech-less right brain hemisphere of a patient to "take a walk." As the patient stands up to go, he then asks the left-brain what it is doing. The left-brain replies with an explanation along the lines of 'oh, I need to get a drink,' to rationalise the behaviour it has found itself performing. In another experiment (1998b, p. 53), both sides of a split brain are shown unrelated images (the right-side a snowy scene, the left a chicken claw) and each asked to choose a relevant matching image. After each side has chosen (a snow shovel and a chicken, respectively) the verbal left-brain was then asked to explain why it chose these two, unrelated images. Without hesitation, it replied that the shovel was to 'clean out the chicken coop.'³

Split-brain confabulation occurs as one-half of a brain rationalises without all of the necessary details available to it. Yet as Dennett (1991, p. 418) argues, the seamlessness of such confabulation demonstrates that such storytelling is as much a part of our normal functioning as web-spinning is for a spider's, and is crucial for our self-identity, a "fundamental tactic of self-protection, self-control, and self-definition" (cf. Zahavi 2005, p. 110). Self here, as should be clear, refers to more than the minimal subjectivity of experience, but to the 'narrative self,' a sense of persisting and evolving identity constructed through embodied habit and episodic memory, through which all reflective experience is interpreted (Schechtman 2011, p. 398). Supporting Dennett's claim, we find many examples of non-pathological or 'broad' confabulation. In Nisbett and Wilson's (1977) famous study, healthy participants were found to overwhelmingly choose the rightmost pair of identical sets of pantyhose as their 'favourite,' and yet to justify their choice based on what they perceived as its aesthetic and tactile qualities. Wheatley (2009, p. 216), on examining these and similar cases—including Libet cases, where subjects' feelings of initiating an action are argued to *follow* its initiation in the brain—argues that 'patient' and 'normal' (that is, pathological and non-pathological) confabulations are differences in degree rather than kind, arguing that the latter "*sound* better because they are based on apriori, causal theories shared by healthy minds." That is to say, what makes pathological confabulation stand-out is not the fact that it occurs, but that it is

unusually disconnected from the personal and communal experience of the confabulator.

Several studies (Ackil and Zaragoza 1998; Ceci et al. 1994) have shown that young children also confabulate, filling in details of remembered events either spontaneously, or in line with the suggestions of experimenters (Ceci et al. 2007). While normal subjects show some susceptibility to suggestion—a phenomenon that shows links with the interpretive faculty involved in confabulation (Gheorghiu and Kruse 1991, pp. 60–61)—the increased suggestibility and confabulation among young children probably stems from the differing rates of development of different memory systems. For example, Tulving (2005, p. 32) has argued that our episodic memory (the 'mental time-travel' of re-living experienced events) develops later (around 4–5 years) than our semantic memory of facts. In one study (Taylor et al. 1994), children were asked to explain how they knew new facts they had learned earlier in the day (in this case, novel colour names like 'chartreuse'). While older children would describe the earlier teaching session, children aged 3–4 years would typically confabulate, claiming, for example, that they had 'always known' the name. Once again, the confabulation is a case of rationalisation without full access to experience, this time due to under-developed episodic memory rather than a severed corpus collusum. Similarly, in cases of dementia, the episodic memory systems on which the narrative self depends deteriorate while the minimal self remains functionally intact (Summa 2014, p. 482). Such patients may have severe deficits in short- or long-term memory, or both, and their confabulation represents a synthesis of their present situation with more ingrained habits and whatever contextualising memories can still be brought to mind (Örülv and Hydén 2006).

The key lesson from these observations is that rationalisation occurs spontaneously and makes use of the pre-conscious information at its disposal, which therefore includes accounting for gaps in that information. The stories we tell ourselves and others to explain our actions are therefore contingent on the information we have access to, and don't necessarily line up with the entire chain of events. The reflective mind imposes reason on its own embodied behaviour. We seem therefore to have an inbuilt tendency to construct and to believe very complex narratives to explain our actions, but our reasoning may be only incidentally or inferentially related to the reasons *why* we do what we do. Sartre's observations, above, show that the 'I'—the sense of being a subject—only appears on reflection, as the central peg on which the narrative is fixed. As Dreyfus (2007a, p. 356) puts it, only when I step back or reflect—either in remembering a past action, or when my smooth-coping is interrupted—can I "then retroactively attach an 'I think' to the coping and take responsibility for my actions."

³ It is rather uncanny that it seems to feel most natural to Gazzaniga and other writers, including myself, to refer to half of a split-brain—which speaks as a conscious person—as an 'it' rather than a 'he' or 'she.'

This account has many parallels with Carruthers (2011) ‘interpretive sensory access’ (ISA) theory of self-knowledge, which he has also applied to explain confabulation (and schizophrenia). Carruthers argues that humans have a single mental faculty underlying our attributions of propositional attitudes, and that we acquire self-knowledge in the same way that we acquire knowledge of others—through the interpretation of sensory data. Self-knowledge is only privileged to the extent that we have access to additional sensory data such as inner speech and proprioception, yet since the interpreter only has access to that data which is present to consciousness, it is not ‘immune’ to error, and frequently infers the wrong source of its actions—that is, it confabulates.

However, my account differs from the ISA in some important ways. Carruthers seems to identify the mind with propositional, reflective consciousness, which takes non-conceptual sensory impressions and forms representations of concepts (which for him include objects, relationships, and so forth), which themselves comprise the input of an interpretive module. Unlike Carruthers, neither Dreyfus nor McDowell endorse a representational theory of mind, and for Dreyfus in particular, pre-reflective experience is not the result of an inner subject representing an outer world, but a being-in-a-world that is already-interpreted in terms of significance. As I have argued above, such interpretations are not primordially conceptual. Our experience while coping is neither of ‘mere’ sensory impressions nor of propositional content, but of meaningful entities and the possibilities they afford—a cake for eating, a frisbee for throwing, a dog for teasing. It is only via breaking out of coping into reflective consciousness that such entities show up *as* objects, that is, as the object of propositional contents.⁴ Since Carruthers does not distinguish between these two modes of experience, he therefore fails to capture the subtle but important distinction between what Dreyfus calls non-conceptual and conceptual experience. A consequence of this, as Sass et al. (2011, p. 9) have noted, is that it seems to locate subjectivity in the created subject of the narrative self, and imply that beings lacking this subject therefore lack awareness of mental states. Thus, where Carruthers locates consciousness *after* the transition into conceptual content, I follow Dreyfus and other phenomenologists in distinguishing two different modes of awareness on either side.

At the heart of my account, then, is a model of the mind involving two distinct forms of content. Firstly, *conceptual* content is tied to explicit, linguistic thought and is *created*

through the process of reflection away from a more primordial intentionality. That second but prior form—the content of our embodied coping—is therefore contrasted as *non-conceptual*. Our brief look at cases of confabulation, both normal and patient, shows that the phenomenon belongs to the former, involving the conceptual, propositional content that is abstracted from prior, non-conceptual enactment. In the next section, I will argue that these two forms of content stem from two modes of cognition, with the latter arising from the interpretation of experience into the construction of the narrative self.

4 Projection and Interpretation

The above phenomenological account is supported by an increasing body of evidence from neurology and the cognitive sciences. A large literature on skill-acquisition (Polanyi 1966; Collins 2000, 2010; Sutton et al. 2011)—that, like Dreyfus, distinguishes between explicit, propositional knowledge and tacit, embodied habitual knowledge—finds neurological parallels in recent work showing that unconscious, habitual skills and actions are associated with the basal ganglia and other mid-brain regions (Graybiel 1998, 2008; Lombo and Giménez-Amaya 2014), while the evaluative monitoring associated with deliberation and problem-solving takes place primarily in the cortex. Hirstein (2005, p. 180), too, traces confabulation principally to the operations of the pre-frontal cortex. This dichotomy is also reflected in the distinction between episodic and semantic memory mentioned above, with the semantic memory of habitual knowledge associated with the mid-brain, while the episodic memory of subjective experience is centred on the pre-frontal cortex (Vandekerckhove et al. 2014). Of particular significance, too, is the fact that both propositional, deliberative content and episodic memory only appear all together in mature human beings, which is consistent with the human cortex being the part of the brain that has evolved most significantly from what we share with non-human animals. Supporting this, Nelson (2003, p. 33) has argued that the narrative self emerges in early childhood as a “new subjective level of conscious awareness” tied to our acquisition of language.⁵ Yet, as Parnas and Sass (2011, p. 525) stress, this richer, reflective, language-based selfhood remains nevertheless based on the “necessary foundation” of a minimal self.

The human brain is, of course, one of the most complex structures in nature, and its many parts do not work in isolation. It would be an oversimplification to say that the pre-frontal and other cortical regions house the rational mind,

⁴ Heidegger (1962, p. 196) distinguishes between the ‘hermeneutic-as’ of our involved practical dealings, and the ‘apophantic-as’ of an object of reflective detachment. See Bergamin (2016, Chap. 3) for an extended discussion of Heidegger’s two ‘as-structures’ and their relation to mental content.

⁵ Thereby situating one of the key motifs of both Heidegger (1993) and McDowell (1994) in developmental psychology.

in opposition to a more primitive, habitual lower brain. The self, consciousness, and rational agency all emerge from the functioning of the brain as a whole. Nevertheless, from the correlations that I and many others have noted—from Kahneman's (2011) 'slow' and 'fast' brains, to the 'predictive processing' model (Clark 2016), to what Davies (2014) calls the 'old' and 'new' brains—there is a growing cognitive scientific consensus around a 'two-systems' model of the mind as comprising two forms of cognition: first, a reflexive mode, centred on the lower and mid-brain, that responds habitually to direct sensory feedback, and largely outside of attentional consciousness; and a second which is rational in the *vernünftig* sense—self-conscious, deliberative, linguistic—and centred on the cortex (cf. Martinez-Valbuena and Bernacer 2014).

Humans are, of course, not alone in our interpretive abilities. There is a sense in which all animal life can be said to 'interpret' the world, insofar as they functionally represent elements of their experience as relevant to their survival (Millikan 1993, pp. 92–93). But 'higher' animals can also interpret the world in terms of other agents and their intentional actions. Chimps, for example, are well-known for their excellent intersubjective awareness. They can follow the gaze of conspecifics and often alter their behaviour accordingly (Tomasello et al. 1998, pp. 1067–1068; Povinelli and Eddy 1996, pp. 133–134), with subordinate chimps, for example, avoiding food that they know a more dominant ape has seen hidden, yet going for it when they know the latter remains ignorant (Hare et al. 2001, p. 144). They also show an awareness of competitors' intentions, and attempt to conceal their actions when there is a possibility of being caught out (Melis et al. 2006, p. 160). Nevertheless, chimp intelligence seems to belong to the habitual, embodied mode of cognition, involving intuitive 'insight' rather than reflective 'foresight' (Seed and Byrne 2010, pp. R1035–R1036; cf.; Köhler 1925). Furthermore, although some animals approach human intelligence in their recognition of objects or kinship relations, it is only in mature human beings that all of these elements come together, and are mediated by the conceptual capacities that create the appropriate, reflective distance for rational thought (McDowell 2007a, p. 346; cf. 1994, pp. 57, 115).

Such animal analogues offer an evolutionary explanation of our tendency to attribute agency to events. An animal's survival is not well-served by wondering whether the rustling in the grass is the wind or a tiger; it is safer to act first and ask questions later, even if the wind rustles far more often than tigers. Along similar lines, it also makes sense for evolution to hard-wire in social responses—it serves a social creature better to swiftly detect who is an ally, who is a threat, and what these friends and foes have in mind, than to lose time logically deducing such relationships. Human beings are also hardwired to be sensitive to these

relationships, as Cosmides and Tooby (1992) have shown, noting what they call a 'cheater detection mechanism' that calls our attention to scenarios where one of our group is acting unfairly (cf. van Lier et al. 2013). Yet such mechanisms are not confined to unconscious or 'intuitive' cognition, but interact with our conceptual reasoning, and in some cases facilitate it. For example, logical problems such as the 'Wason Selection Task' (Wason 1966) are much more quickly and intuitively solved when framed in terms of social contracts than when presented as an abstract logic puzzle (Cosmides and Tooby 1992, p. 184).

If animals can directly perceive agency, and social animals can directly perceive others' intentions, then we should not be surprised that complex socio-cultural animals like ourselves project intentional agency and mindedness onto the world. And this is indeed what we see, in our relationships not only with ourselves, but in the ways we deal with nature and with technology. We say that the dog is jealous of the cat, which may or may not be true, but also that our iPod is jealous of our iPhone, which certainly is not (Luhrmann 2010). We talk about our computers as 'thinking' and our cars as 'tired.' A storm sailing over the plains may feel purposive, seeking to hunt us down as we cycle frantically towards the village. Indeed, Piaget (1929) related animist and nature religions to an innate tendency to project agency onto natural objects, such as the weather, fire, and plants and animals, while Abram (1996, p. 303) has more recently suggested that reflective rationality grows out of an embodied knowledge that projects the action of mind onto perceived objects and events.

Therefore, I argue, the human experience of agency both extends and diverges from that of other animals along the lines of the two cognitive systems I described in the previous sections. To the extent that we share habitual, non-conceptual embodied cognition with other animals, we possess the capacity to directly *perceive* elements of the world as having agency. Yet we also, through the exclusively-human conceptual cognition centred in our fully-developed cortices, *interpret* actions as the rational acts of other agents. This extends from natural events, like storms, to artefacts, like computers, as well as to the acts of conscious beings—including ourselves. On this account, then, confabulation is tied to *self*-consciousness, and is the rational, conceptual interpretation of our own spontaneous, non-conceptual agency.

5 Schizotypy as Over-interpretation

I have so far been arguing that confabulation, as a mismatch between our interpretative faculties and the source of our agency, is a symptom not only of pathological cross-wiring, but also of ordinary sense-making, revealing the partially-contingent nature of the interpretation of our primordial

experience into propositional, conceptual content. In this final section, I will explore the consequences of this account for mental health, suggesting that the narratives formed by our interpretive faculties are essential for a stable sense of self. However, I will argue that a hyperactivity of interpretation results in a schizotypal personality, leading in extreme cases to schizophrenia.

Philosophers such as Dennett (1991, 1992), Taylor (1989), and Schechtman (2011) argue that our personal identity and consciousness of ourselves *as selves* are dependent on the stories that we (and others) tell, which create a coherent (and shared) narrative of an agent acting over time. Such accounts, I have suggested, are consistent with, and dependent on, the interpretive faculty I have been arguing for, with its creation of propositional content from immersed experience in the act of rationalising the content (and gaps) of memory. These narrative selves are, to a large degree, a self-reinforcing loop, as decisions are justified to fit the narrative, which is itself constructed as the sum of past choices and events.

This process plays an important role in maintaining robust mental health. In its spontaneous operation, it reinforces a sense of empowerment and agency, as in Nisbett and Wilson's (1977) study mentioned above, where a choice that was essentially meaningless (since the pairs of pantyhose were identical) became a vehicle for the chooser's values (good quality, and so on). The same principle can, over time, encourage ownership of poor or potentially-regrettable decisions, by building them into a narrative where they were 'meant-to-be' or led ultimately to positive outcomes. Carried to an extreme, such a tendency may become a hubristic belief that one's decisions never err. But in moderation, the ability to own and justify one's choices leads on the one hand to the acceptance of misfortune and growth through incorporating it into one's life-story, while on the other hand avoiding a Hamlet-like paralysis in the face of challenging dilemmas (cf. Frankl 2004). Yet even this positive aspect involves the alteration of memories over time, as reasons are justified and stories repeated, slowly diverging from an original experience that may itself have been unconsciously confabulated.

Such confabulation is natural, and while unconscious, it nevertheless makes use of contextual cues that guide the subject's interpretation of their experience, in what might be described as a form of inferential, albeit intuitive, reasoning (Örülv and Hydén 2006, p. 653; cf. Carruthers 2011, p. 328). Similarly, since normally-functioning minds are 'wired' to detect patterns and agency, 'illusory pattern perception'—the detection of pattern-like sequences in random stimuli—is surprisingly common (Williams and Griffiths 2013). 'Schizotypal' minds are particularly prone to over-interpreting situations to see patterns which aren't there (Davies 2014, p. 91). Coincidences take on additional "salience" (Kapur 2003,

pp. 14–15) and random events are incorporated into a larger meta-narrative that makes sense of the world, although in ways that diverge significantly not only from the prevailing cultural narrative, but from what the subject could reasonably deduce from their experience.

'Schizotypy' refers to a "latent personality organisation" that has many of the symptoms of schizophrenia, including noticeable behavioural, emotional, and perceptual abnormalities, but falls short of the more intense, and frequently psychotic, dissociation that marks full-blown schizophrenia (Ettinger et al. 2015). 'Latency' here indicates a "latent liability" to developing schizophrenia, although the path of development includes a complex mix of neurological and cultural-environmental factors, and there is no single determining factor that predicts whether a schizotypal personality remains sub-clinical or develops more serious schizophrenic symptoms (Lenzenweger 2015).

On the neurological level, until very recently the leading neurological theory of schizophrenia placed a central role on the overproduction of dopamine (Davies 2014, p. 91), although it should be noted that the neurotransmitter's exact role in schizophrenia is now the subject of serious debate, and neuroscience is still not in a position to say what, if any, causal role it plays in the precise mix of chemical and environmental causes (Kendler and Schaffner 2011, p. 59). There remains, nevertheless, a correlation between excessive dopamine and schizophrenic symptoms. Schizophrenics, during psychotic episodes, show a heightened synthesis of dopamine, and a higher release of it in response to impulses, while dopamine-stimulators such as amphetamines can produce or amplify acute psychotic effects in schizotypal personalities (Kapur 2003, pp. 14–15). Negative correlations further support this picture, with anti-psychotic medications functioning by blocking dopamine receptors, and thus preventing the overactive pattern-recognition and reinforcement that would otherwise occur (*ibid*, p. 16).

Dopamine is found in all multicellular animals, and plays an important role in the learning and reward system, where its release accompanies and reinforces successful actions and pattern recognition (Barron et al. 2010). Specifically, its release accompanies behaviour that it benefits the animal to repeat. These tend to be attraction or avoidance, and so dopamine has often been described as both reward and punishment (*ibid*), although contrary to popular misconception, the dopamine itself is not the reward—it is not a 'feel-good' chemical like endorphins. Its function is rather to stimulate the repetition of an action, and to increase the expectation of a reward.

Dopamine therefore plays an important role in habit formation, and is crucial to the operation of the non-conceptual skill-acquisition discussed above (Graybiel 2008, pp. 370–371). Davies (2014, pp. 91–92) links this process with pattern-recognition and argues that, phenomenologically, we

are confronted with patterns first as feelings, which are then rationalised as they rise into propositional consciousness—an account which is consistent with my understanding of conceptual consciousness as arising through interpretation. Thus, it is not the presence of excessive dopamine *per se* which is the issue here, but rather its effect on the operations of the cortically-centred interpretive faculty, since, unlike many other psychiatric disorders, schizophrenia appears to have no analogues in other animals (Ogawa and Vallender 2014). Schizophrenia and schizotypy might therefore arise as a disorder of the human, conceptually cognitive faculty of interpretation, and indeed, some researchers have suggested that schizophrenia is a “maladaptive by-product of changes in the brain that led to the human-specific traits of language, social cognition and interpersonal behaviour” (*ibid*; cf. Crow 1997; Burns 2004; Horrobin 1998).

I therefore suggest that we understand schizotypal-spectrum disorders as characterised by an excess of interpretation, detecting patterns in the ‘random noise’ of ordinary sensory and emotional stimulation. Such over-interpretation would also account for the more sensory symptoms of schizophrenic psychosis, such as hallucination, which Kapur (2003, p. 15) argues are “exaggerated, amplified, and aberrantly recognised internal percepts” (cf. Grossberg 2000; Bentall 1990). For example, our cortex contains specific modules that interpret faces, language, and nefarious actions such as cheating (Kanwisher and Dilks 2013; Cosmides and Tooby 1994). Overactive interpretation therefore explains such hallucinations as seeing faces in complex visual input like clouds or wallpaper, or hearing voices in ambient noise, as well as paranoid symptoms like inferring malevolent intentions in other people.

Similarly, Kapur (2003, p. 15) argues that schizotypal delusions are “essentially disorders of inferential logic, as most delusional beliefs are not impossible, just highly improbable.” That is to say, while such delusions may be far-fetched, they are often held together with an internal logic that puts them on a spectrum with ordinary world-views. Indeed, logically-consistent but unjustified beliefs are as widespread today as they have been throughout history, as witnessed by the everyday superstitions and ritualised routines practised even by professional rationalists. These, I have suggested, are only slightly distanced from the everyday confabulation that we perform to interpret and give meaning to our actions.

A further correlation between interpretation, confabulation, and schizophrenia is supported by the convincing links that Barron et al. (2018) have made in multiple studies connecting schizotypy to belief in conspiracy theories and magical thinking. Magical thinking is not necessarily pathological, being an almost ubiquitous feature of pre-industrial societies, and which is often not, strictly-speaking, irrational, since it may, as I suggested earlier, stem directly

from interpreting natural events as enactments of agency. However, van Prooijen et al. (2018) have drawn connections between supernatural beliefs in modern societies, belief in conspiracy theories, and the ‘illusory pattern perception’ that is a key feature of schizotypy, while Whitson and Galinsky (2008) have demonstrated links between ‘illusory pattern perception’ and feelings of a lack of control. The formulation of and belief in conspiracy theories—the “belief that the world or an event is manipulated by omnipresent and omnipotent agents in the pursuit of malevolent goals” (Barron et al. 2018, p. 15)—takes such thinking to an extreme, where a single force (the government, the ‘deep-state,’ the Illuminati, the Jews—themselves all understood as single-minded actors) is taken to be responsible for a range of unrelated events, both on the macro-scale (political and economic events), and the micro-scale of individual lives (surveillance, manipulation, and so on).⁶ None of these links imply that ordinary ‘illusory pattern perception’ leads to schizotypy, much less schizophrenia; rather, they support the hypothesis that these (latent or overt) pathologies lie at the extreme end of a spectrum that begins with our ordinary interpretive faculties.

Kapur (2003, p. 13) has noted that the content of schizophrenic delusions is culture-specific—a South African villager might attribute perceived persecution to an evil shaman, while a Canadian urbanite might interpret similar feelings as the machinations of the Mounties. A religious voice-hearer might hear angels, while an atheist receives messages from aliens. This suggests that the inferences of schizotypes are drawn from the extensive web of beliefs that they have both inherited from their culture, and established over a lifetime. Conspiracy theorists are notoriously difficult to argue with, not only because their beliefs tend to be unfalsifiable, but because their belief schema—the meta-narrative of conspiracy—provides a ready framework for the inference and incorporation of any new facts. Yet this process of interpretation is not, of itself, pathological, but is rather an extension or over-reaction of the everyday confabulation that causes subjects not only to find reasons for preferring one pair of pantyhose over an identical one, for example, but to genuinely believe that they do.

The parallels to confabulation indicate that, although schizotypal and psychotic delusions, conspiracy theories, and magical thinking are all inferences drawn between perceived patterns within a particular cultural milieu, they are not necessarily the result of conscious reasoning. Rather, they appear in consciousness—as propositional, conceptual

⁶ The degree to which a conspiracy theorist feels personally, rather than abstractly, targeted by such forces seems to correlate to states of pathological paranoia rather than latent schizotypy, and may warrant further investigation.

content—*through* the process of interpretation of primordial, non-conceptual experience, within the context of an existing belief schema acquired through enculturation and repeated practice. Phenomenological accounts support this, with Parnas and Sass (2011, p. 537) arguing that schizotypy and schizophrenia are characterised by what they call “hyper-reflexivity,” where objects which would ordinarily form part of the non-conceptual background are forced into conceptual consciousness. Less salient elements—including features of the perceptual environment along with kinaesthetic bodily sensations, ‘inner speech,’ and the processes or presuppositions of thinking—become elements of experience in their own right, which must be accounted for as part of the subject’s immediate and longer-term understanding. Just as a dementia patient confabulates to make sense of the living room she doesn’t recognise, the schizophrenic needs to find an explanation of the thoughts he doesn’t identify with. Sass et al. (2011, p. 9) have argued that this hyper-reflexivity is a disorder of the minimal self, but I suggest that the issue should be situated rather in the *transition*—that is, during the interpretation of pre-reflective, non-conceptual experience into the conceptual content that becomes the object to the subject of the narrative self. The minimal self, as an underlying subjectivity, is still present prior to this transition, at least as the shadow of the not-self that is experienced as intruding. It may be possible that, in extreme cases, the entire process and even the core self break down completely. In such cases, it is difficult to imagine phenomenologically what such a breakdown would be like, or to conceive of a consciousness without *any* sense of subjectivity, not even a subjectivity disconnected from all sense of its own agency. But here, perhaps, we start to approach what Jaspers (1997, pp. 27, 98) described as the “un-understandability” of psychosis.

For this very reason, Parnas and Sass (2001, p. 102) recommend investigations into the phenomenology of schizophrenia begin with schizotypy (cf. Zahavi 2005, pp. 133–134). And through exploring the links between confabulation and schizotypic thinking, I have suggested that the root issue lies neither with the minimal nor narrative selves *per se*, but with the interpretation of content that creates the latter from the former. The spontaneous and unconscious rationalisation, revealed in confabulation, that powers this interpretive faculty also hints at the difficulties in dealing both with paranoid delusions, as well as irrational but non-psychotic conspiratorial thinking. Kapur has stressed that there is no ‘overnight cure’ to schizophrenia, since complete recovery requires the patient to process and change their belief schema, and the way in which they understand themselves and their relation to the world. Anti-psychotic drugs provide only symptomatic control; they “do not primarily change thoughts or ideas” but rather “provide a neurochemical milieu in which new aberrant saliences are less

likely to form and previously aberrant saliences are more likely to extinguish” (Kapur 2003, p. 17). That is to say, they moderate the hyperactivity of the interpretative faculty—and hence stop the over-interpretation of patterns—but they are unable on their own to affect the belief schema through which that faculty operates. However, by focusing on the ‘belief schema,’ Kapur risks overlooking how the problem is situated in *interpretation* itself. Treatment, therefore, cannot simply involve (the often painstaking) persuasion of a patient to conceptually adjust their meta-narrative, but must also focus on the pre-reflective interpretive process that *creates* the very content of that narrative.

6 Conclusions

In this paper, I have argued that our sense of self as subject arises from an interpretive faculty that creates conceptual content from our immediate experience, incorporating it into the narrative of our life-story and inherited cultural beliefs. This faculty is above all a pattern-detector, and wired to perceive certain elements, such as faces, language, and—importantly, in terms of this special issue—agency. The phenomenon of confabulation occurs when our rational mind interprets its own pre-reflective agency through the lens of the social and cultural expectations in which it is immersed, with a tendency to interpret its actions in ways that give us agency and ownership, even on occasions when this is not fully justified. I have suggested that this plays a role in maintaining robust mental health. However, I have argued that excessive or hyperactive interpretation results in schizotypal-spectrum disorders. Schizophrenia, I have suggested, therefore has its roots in the over-interpretation and misapplication of conceptual content.

I argued this by first fleshing out the distinction between conceptual and non-conceptual content through a discussion of the Dreyfus-McDowell debate. I argued that Dreyfus’ phenomenology reveals a distinction between our enactments of embodied, reflexive skill, and the reflection on such enactments that rationalises it into a propositional content. I argued that McDowell’s criticisms did not refute such a distinction, but helped to refine our understanding of rationality not as an expression of ‘reasons why’ (*Gründe*) as such, but as a faculty of linguistic, reflective distance (*Vernunft*). I followed Carman in holding that the propositional content of this faculty is not always present, but is actively *created* in the process of reflective rationalisation. I argued that cases of non-pathological confabulation demonstrate this process and the resulting gap between experience and reflection, and that more severe cases are the result of an impaired access to the originary experience. I then drew upon neurological and psychological evidence to locate this interpretive faculty in the higher, cortical brain regions in which, together with the

episodic memory on which it relies to function correctly, it belongs to a cluster of capacities that separate human beings from non-human animals.

Thus, I argued, confabulation is a side-effect of the normal functioning of our rational, interpretive faculty. In ordinary cases, this faculty combines our experience with our expectations—both personal and cultural—in order to give meaning to our actions, confabulating to fill in or pass over gaps and inconsistencies in those combinations. This faculty aims, above all, at coherence and consistency between our experience and the meta-narrative or ‘belief schema’ through which it operates, and which it is geared to prioritise over what we might call objective reality. Therefore, in cultures with a large place given over to magical thinking, the agency of supernatural forces finds a greater role in interpretation.

In all cultures, however, there is a small subset of people who *over-interpret* their direct perceptions. Schizotypal personalities—due, I have suggested, at least in part to the effects of the excessive action of cortical dopamine—perceive cause, effect, and agency in coincidental events, and detect patterns in random noise. In extreme or psychotic cases of sensory over-interpretation, this leads to hallucinations; in cases of agency, it leads to delusions and the perception of powerful supernatural forces, as well as unrealistic, confabulated beliefs about the self—identifying with a role or even a personality that is not one’s own. Like ordinary confabulation, these interpretations are mediated through the prevailing culture, and in some sense are a rational integration of over-interpreted experience with a distorted meta-narrative—with a resulting feedback loop that increases dissociation from public, communal narratives. Schizophrenia, therefore, ought to be thought of as a disorder of that interpretive faculty which gives rise to our narrative, rational concept of self.

My account is an empirical one, combining philosophical concepts with evidence from neuropsychology, and the links I have suggested should continue to be testable as our knowledge of the brain improves. Yet its implications are also cultural, for I have suggested throughout that the activity of the embodied brain is not separable from its cultural milieu. Much work has already been done on the ways different understandings of psychosis in pre-industrial societies affect integration and recovery (e.g., Peters 1982). But the more interesting questions, to my mind, arise from the fragmenting shared meta-narrative of our post-industrial society, and particularly the increasing influence of conspiracy theories in the digital age (which themselves spread through the dopamine-stimulating networks of the web and social media). If schizophrenia is an extreme example of the confabulatory interpretation that marries our experience with our personal and cultural narratives, what are the implications for ourselves as interpreters when those narratives are in flux?

Acknowledgements My thanks to two anonymous reviewers for their challenging but encouraging criticism, which greatly improved this paper, and also to colleagues at Durham University for the many conversations that helped shape the theoretical background. The research for this paper was funded, for the most part, by a doctoral award from Durham University, but special thanks are due to River MacDonald for her generous hospitality in West Lothian, which provided an ideal retreat for the writing of this paper.

Funding The research for this study was partially-funded by Durham University (Durham Doctoral Award).

Compliance with Ethical Standards

Conflict of interest The author declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by the author.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abram D (1996) *The spell of the sensuous*. Pantheon, New York
- Ackil JK, Zaragoza MS (1998) Memorial consequences of forced confabulation: age differences in susceptibility to forced memories. *Dev Psychol* 34:1358–1372
- Baars BJ (2002) The conscious access hypothesis: origins and recent evidence. *Trends Cogn Sci* 6:47–52
- Barron AB, Søvik E, Cornish JL (2010) The roles of dopamine and related compounds in reward-seeking behavior across animal phyla. *Front Behav Neurosci* 4:163
- Barron D, Furnham A, Weis L, Morgan KD, Towell T, Swami V (2018) The relationship between schizotypal facets and conspiracist beliefs via cognitive processes. *Psychiatry Res* 259:15–20
- Bentall RP (1990) The illusion of reality: a review and integration of psychological research on hallucinations. *Psychol Bull* 107:82–95
- Bergamin JA (2016) *In the beginning was the word: concepts, perception, and human being*. Dissertation, University of Durham
- Burns JK (2004) An evolutionary theory of schizophrenia: cortical connectivity, metarepresentation, and the social brain. *Behav Brain Sci* 27:831–855
- Carman T (2013) Conceptualism and the scholastic fallacy. In: Scheer JK (ed) *Mind, reason, and being-in-the-world: the McDowell-Dreyfus debate*. Routledge, Abingdon
- Carruthers P (2011) *The opacity of mind: an integrative theory of self-knowledge*. Oxford University Press, Oxford
- Ceci SJ, Huffman MLC, Smith E, Loftus EF (1994) Repeatedly thinking about a non-event: source misattributions among preschoolers. *Conscious Cogn* 3:388–407
- Ceci SJ, Kulkofsky S, Klemfuss JZ, Sweeney CD, Bruck M (2007) Unwarranted assumptions about children’s testimonial accuracy. *Annu Rev Clinical Psycho* 3:311–328
- Clark A (2016) *Surfing uncertainty: prediction, action, and the embodied mind*. Oxford University Press, Oxford

- Collins HM (2000) Four kinds of knowledge, two (or maybe three) kinds of embodiment, and the question of artificial intelligence. In: Wrathall MA, Malpas J (eds) Heidegger, coping, and cognitive science: essays in honor of Hubert L. Dreyfus, vol 2. MIT Press, Cambridge
- Collins HM (2010) Tacit and explicit knowledge. University of Chicago Press, Chicago
- Cosmides L, Tooby J (1992) Cognitive adaptations for social exchange. In: Markow J, Cosmides L, Tooby J (eds) The adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, Oxford
- Cosmides L, Tooby J (1994) Beyond intuition and instinct blindness: toward an evolutionarily rigorous cognitive science. *Cognition* 50:41–77
- Crane T (2013) The given. In: Schear JK (ed) Mind, reason, and being-in-the-world: the McDowell-Dreyfus debate. Routledge, Abingdon
- Crow TJ (1997) Is schizophrenia the price that Homo sapiens pays for language? *Schizophr Res* 28:127–141
- Davies J (2014) Riveted. Palgrave Macmillan, New York
- Dennett DC (1991) Consciousness explained. Little, Brown & Co, Boston
- Dennett DC (1992) The self as a center of narrative gravity. In: Kessle F, Cole P, Johnson D (eds) Consciousness and self: multiple perspectives. Laurence Erlbaum, Hillsdale
- Dreyfus HL (2005) Overcoming the myth of the mental: how philosophers can profit from the phenomenology of everyday expertise. In: Proceedings and addresses of the American Philosophical Association, vol 79, pp. 47–65
- Dreyfus HL (2007a) The return of the myth of the mental. *Inquiry* 50:352–365
- Dreyfus HL (2007b) Response to McDowell. *Inquiry* 50:371–377
- Dreyfus HL (2013) The myth of the pervasiveness of the mental. In: Schear JK (ed) Mind, reason, and being-in-the-world: the McDowell-Dreyfus debate. Routledge, Abingdon
- Dreyfus HL, Dreyfus SE (1986) Mind over machine. Blackwell, Oxford
- Ettinger U, Mohr C, Gooding DC, Cohen AS, Rapp A, Haenschel C, Park S (2015) Cognition and brain function in schizotypy: a selective review. *Schizophrenia Bull* 41:S417–S426
- Frankl VE (2004) Man's search for meaning. Lasch I (trans). Rider, London
- Gazzaniga MS (1998a) The mind's past. University of California Press, Berkeley
- Gazzaniga MS (1998b) The split brain revisited. *Sci Am* 239:50–55
- Gazzaniga MS, LeDoux JE, Wilson DH (1977) Language, praxis, and the right hemisphere: clues to some mechanisms of consciousness. *Neurology* 27:1144–1147
- Gheorghiu V, Kruse P (1991) The psychology of suggestion: an integrative perspective. In: Schumaker JF (ed) Human suggestibility. Routledge, New York
- Gibson JJ (1979) The ecological approach to visual perception. Houghton Mifflin Co, Boston
- Graybiel AM (1998) The basal ganglia and chunking of action repertoires. *Neurobiology Learning Memory* 70:119–136
- Graybiel AM (2008) Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* 31:359–387
- Grossberg S (2000) How hallucinations may arise from brain mechanisms of learning, attention, and volition. *J Int Neuropsych Soc* 6:583–592
- Hare B, Call J, Tomasello M (2001) Do chimpanzees know what conspecifics know? *Anim Behav* 61:139–151
- Heidegger M (1962) Being and time. Macquarrie J, Robinson E (trans). SCM Press, London
- Heidegger M (1993) Letter on humanism. Capuzzi FA (trans). In: Krell DF (ed) Basic Writings. Routledge, London
- Hirstein W (2005) Brain fiction: self-deception and the riddle of confabulation. MIT Press, Cambridge
- Horrobin DF (1998) Schizophrenia: the illness that made us human. *Med Hypotheses* 50:269–288
- Jaspers K (1997) General psychopathology. Hoenig J, Hamilton MW (trans). Johns Hopkins University Press, Baltimore
- Kahneman D (2011) Thinking, fast and slow. Farrar, Straus and Giroux, New York
- Kanwisher N, Dilks D (2013) The functional organization of the ventral visual pathway in humans. In: Chalupa L, Werner J (eds) The new visual neurosciences. Bradford/MIT Press, Cambridge
- Kapur S (2003) Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry* 160:13–23
- Kendler KS, Schaffner KF (2011) The dopamine hypothesis of schizophrenia: an historical and philosophical analysis. *Philos Psychiatry Psychol* 18:41–63
- Köhler W (1925) The mentality of Apes. Winter E (trans). Kegan Paul, London
- Lenzenweger MF (2015) Thinking clearly about schizotypy: hewing to the schizophrenia liability core, considering interesting tangents, and avoiding conceptual quicksand. *Schizophrenia Bull* 41:S483–S491
- Lombo JA, Giménez-Amaya JM (2014) The unity and the stability of human behavior. An interdisciplinary approach to habits between philosophy and neuroscience. *Front Hum Neurosci* 8:607
- Luhrmann T (2010) What students can teach us about iPhones. Salon.com, May 30th 2010
- Martinez-Valbuena I, Bernacer J (2014) Behavioral duality in an integrated agent. *Front Hum Neurosci* 8:614
- McDowell J (1994) Mind and world. Harvard University Press, Cambridge
- McDowell J (2007a) What myth? *Inquiry* 50:338–351
- McDowell J (2007b) Response to Dreyfus. *Inquiry* 50:366–370
- McDowell J (2013) The myth of mind as detached. In: Schear JK (ed) Mind, reason, and being-in-the-world: the McDowell-Dreyfus debate. Routledge, Abingdon
- Melis AP, Call J, Tomasello M (2006) Chimpanzees conceal visual and auditory information from others. *J Comp Psychol* 120:154–162
- Millikan RG (1993) Biosemantics. In: White queen psychology and other essays for Alice. Bradford/MIT Press, Cambridge
- Nelson K (2003) Narrative and the emergence of a consciousness of self. In: Fireman GD, McVay TE, Flanagan OJ (eds) Narrative and consciousness: literature, psychology and the brain. Oxford University Press, Oxford
- Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. *Psychol Rev* 84:231–259
- Ogawa LM, Vallender EJ (2014) Evolutionary conservation in genes underlying human psychiatric disorders. *Front Hum Neurosci* 8:283
- Okrent M (2007) Rational animals. Ohio University Press, Athens
- O'Regan JK (2000) Experience is not something we feel but something we do. Talk at ASSC 2000, quoted from Tye M (2006) Nonconceptual content, richness, and fineness of grain. In: Gendler TS, Hawthorne J (eds) Perceptual Experience. Oxford University Press, Oxford.
- Örülü L, Hydén LC (2006) Confabulation: sense-making, self-making and world-making in dementia. *Discourse Stud* 8(5):647–673
- Parnas J, Sass LA (2001) Self, solipsism, and schizophrenic delusions. *Philos Psychiatry Psychol* 8:101–120
- Parnas J, Sass LA (2011) The structure of self-consciousness in schizophrenia. In: Gallagher S (ed) The Oxford handbook of the self. Oxford University Press, Oxford
- Peters LG (1982) Trance, initiation & Psychotherapy in Tamang Shamanism. *Am Ethnol* 9:21–46

- Piaget J (1929) The child's conception of the world. Kegan Paul, London
- Polanyi M (1966) The logic of tacit interference. *Philosophy* 41:1–18
- Povinelli DJ, Eddy TJ (1996) Chimpanzees: joint visual attention. *Psychol Sci* 7:129–135
- Rietveld E (2010) McDowell and Dreyfus on unreflective action. *Inquiry* 53(2):183–207
- Rouse J (2013) What is conceptually articulated understanding? In: Schear JK (ed) *Mind, reason, and being-in-the-world: the McDowell-Dreyfus debate*. Routledge, Abingdon
- Sandifer PH (1946) Anosognosia and disorders of the body schema. *Brain* 69:122–137
- Sartre J-P (2004) *The transcendence of the Ego*. Brown A (trans). Routledge, Abingdon
- Sass L, Parnas J, Zahavi D (2011) Phenomenological psychopathology and schizophrenia: contemporary approaches and misunderstandings. *Philos Psychiatry Psychol* 18:1–23
- Schear JK (2013) Are we essentially rational animals? In: Schear JK (ed) *Mind, reason, and being-in-the-world: the McDowell-Dreyfus debate*. Routledge, Abingdon
- Schechtman M (2011) The narrative self. In: Gallagher S (ed) *The oxford handbook of the self*. Oxford University Press, Oxford
- Seed A, Byrne R (2010) Animal tool-use. *Curr Biol* 20:R1032–R1039
- Siegel S (2016) The contents of perception. In: Zalta EN (ed) *The stanford encyclopedia of philosophy* (Spring 2016 Edition). <http://plato.stanford.edu/archives/spr2016/entries/perception-contents/>
- Summa M (2014) The disoriented self. Layers and dynamics of self-experience in dementia and schizophrenia. *Phenomenol Cogn Sci* 13(3):477–496
- Sutton J, McIlwain D, Christensen W, Geeves A (2011) Applying intelligence to the reflexes: embodied skills and habits between Dreyfus and Descartes. *J Brit Soc Phenomen* 42:78–103
- Taylor C (1989) *Sources of the self*. Cambridge University Press, Cambridge
- Taylor C (2002) Foundationalism and the inner-outer distinction. In: Smith NH (ed) *Reading McDowell: on mind and world*. Routledge, London
- Taylor M, Esbensen BM, Bennett RT (1994) Children's understanding of knowledge acquisition: the tendency for children to report that they have always known what they have just learned. *Child Dev* 65:1581–1604
- Tomasello M, Call J, Hare B (1998) Five primate species follow the visual gaze of conspecifics. *Anim Behav* 55:1063–1069
- Tulving E (2005) Episodic memory and autonoesis: uniquely human? In: Terrace HS, Metcalfe J (eds) *The missing link in cognition: origins of self-reflective consciousness*. Oxford University Press, Oxford
- Van Lier J, Revlin R, De Neys W (2013) Detecting cheaters with thinking: testing the automaticity of the cheater detection module. *PLoS ONE* 8:1
- Van Prooijen J-W, Douglas KM, De Inocencio C (2018) Connecting the dots: Illusory pattern perception predicts belief in conspiracies and the supernatural. *Eur J Soc Psychol* 48:320–335
- Vandekerckhove M, Bulnes LC, Panksepp J (2014) The emergence of primary anoetic consciousness in episodic memory. *Front Behav Neurosci* 7:210
- Wason P (1966) Reasoning. In: Foss BM (ed) *New horizons in psychology*. Penguin, Harmondsworth
- Wheatley T (2009) Everyday confabulation. In: Hirstein W (ed) *Confabulation: views from neuroscience, psychiatry, psychology, and philosophy*. Oxford University Press, Oxford
- Whitson JA, Galinsky AD (2008) Lacking control increases illusory pattern perception. *Science* 322:115–117
- Williams JJ, Griffiths TL (2013) Why are people bad at detecting randomness? A statistical argument. *J Exp Psychol Learn* 39:1473–1490
- Zahavi D (2005) *Subjectivity and selfhood*. Bradford/MIT Press, Cambridge, MA